The role of cross-linguistic lexical similarity on bilingual word acquisition

Gonzalo García-Castro, Daniela Avila-Varela, Núria Sebastian-Galles

SAP seminar

September 22nd 2020

Bilingual lexical acquisition

One-to-three years old bilinguals show smaller **conceptual** vocabulary sizes than monolinguals, but similar **total vocabulary** sizes (e.g., Core et al. 2013).

This suggests that **translation equivalents** (TEs) play a central role during this stage of lexical acquisition.

At 24mo, bilinguals learning languages that share many **cognates** (form-similar TEs) show **larger vocabularies in their additional language**. This effect is larger for productive vocabulary (Floccia et al., 2018).

Parallel activation

Language non-selective account: bilinguals adults (e.g., Costa et al., 2000) and toddlers (Von Holzen & Mani, 2012) **activate words in both languages** even when only using one of them.

Translation equivalents are activated in **parallel**, which impacts the performance of bilinguals in both comprehension and production tasks.How does parallel activation (an on-line phenomenon) impact **lexical acquisition** (an off-line event)?

Anchoring networks: Words sharing high **semantic** (Hills et al., 2009) and/or **phonological** (Fourtassi et al., 2020) similarity with structured **networks** of aquired words are more likely to be acquired next.

Our study

If **parallel activation** plays a role in lexical acquisition:

Words in one language should aid the acquisition of words in the other.

1) Cognate TEs should be acquired **earlier**
2) Cognate pairs of TEs should be acquired **closer in time**
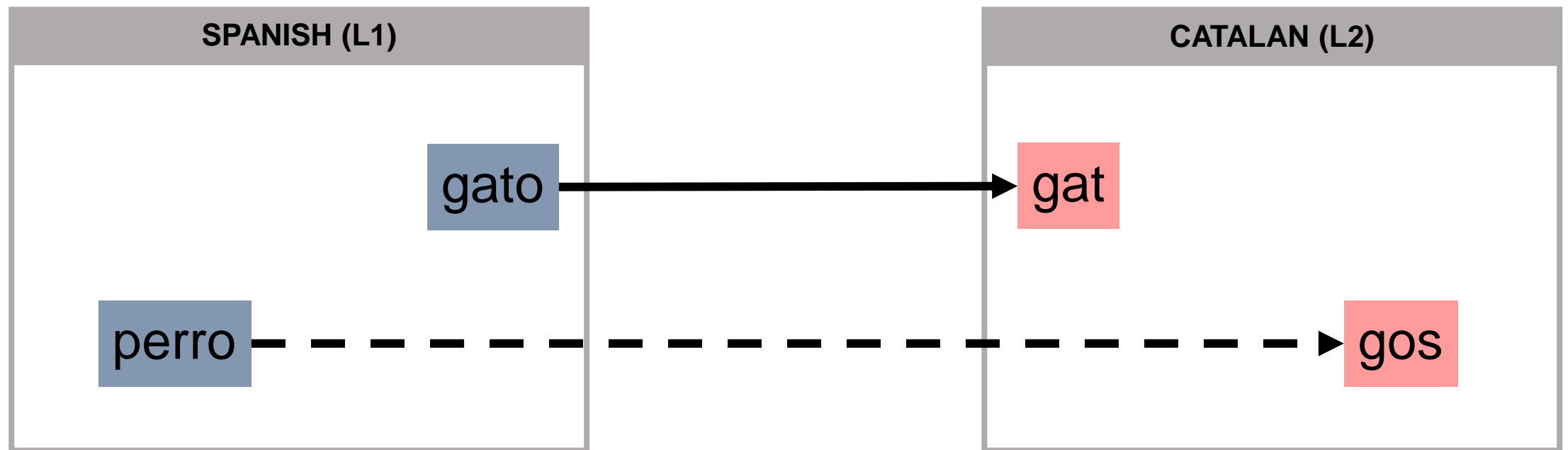
Vocabulary checklist

BiLexicon [alternative names wanted]
   "mesa": Understands, Understands and Says, None

On-line (~200/800 items per language) or lab based (~100 items per language)

16 functional/semantic **categories** (e.g., animals, adverbs)

Included in analyses: **nouns**, **verbs** and **adjectives** (Fourtassi, 2020)

Data from 696 **translation equivalents** (526 cognates, 172 non-cognates)

Vocabulary checklist

Lexical frequency: a potential confound (cognates could be more frequent)

Lexical frequencies from SUBTLEX-ESP (Cuetos et al., 2011) and SUBTLEX-CAT (Boada et al., 2019). Based on TV show subtitles (adult norms)

No differences in lexical frequency across:

- Languages (BF_{01} = 12.26 ± 181 0.001%)
- Cognates status (BF_{01} = 14.39 ± 0.000%)
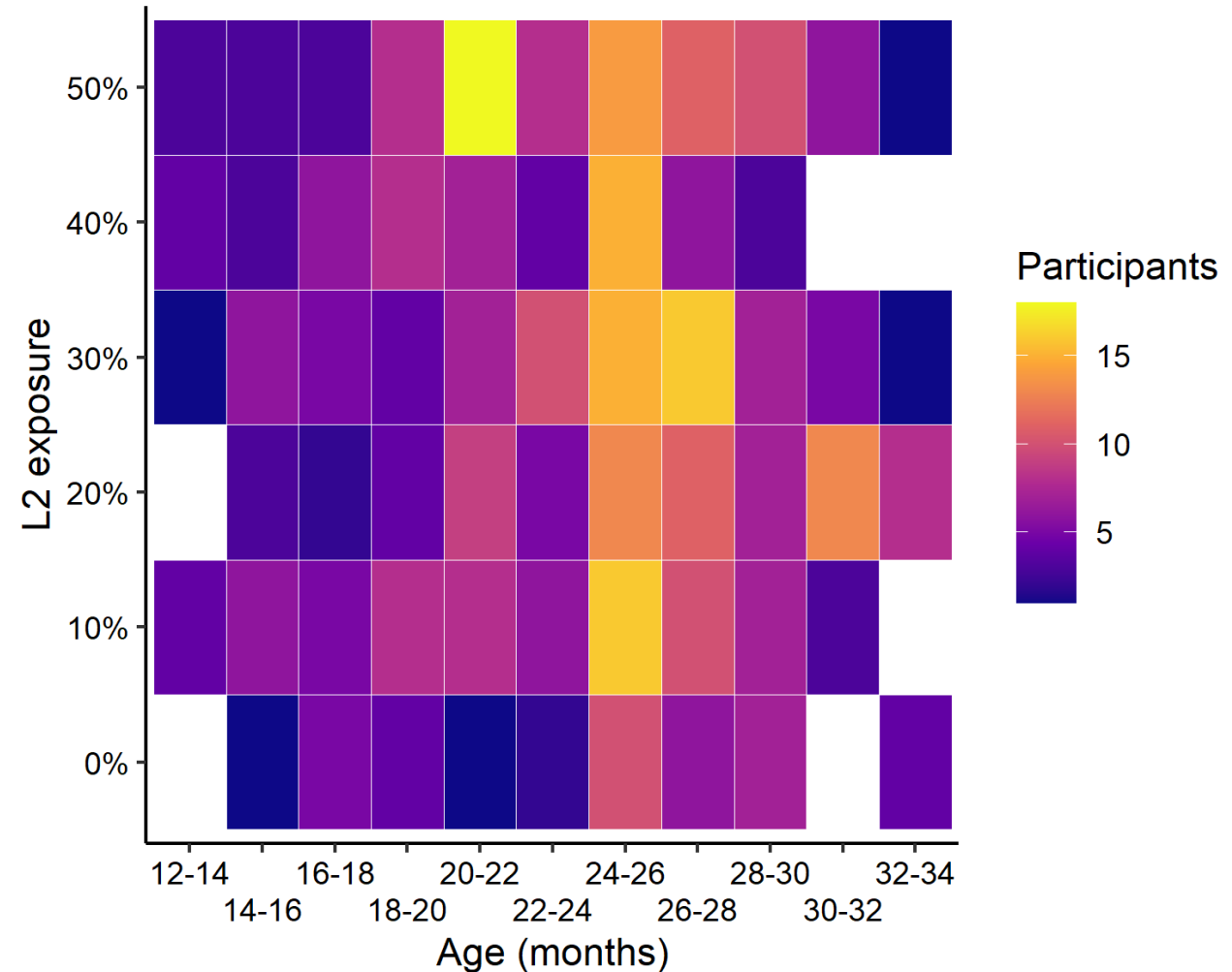- Their combinations (BF_{01} = 699.65 ± 0.016%).

Sample

N = **437**, bilinguals (**12-34 mo**)

Exposure to Spanish and Catalan (%) as reported by parents

Participants considered **bilinguals** if exposed to 5% of a 2nd language

Excluded participants with >10% exposure to a 3rd language.

Design

Dominance: L1 (language of most exposure) vs. L2

Item "taula" is L1 for Catalan-dominant participants

Item "mesa" is L1 for Spanish-dominant participants

Cognateness: Cognate vs. Non-cognate

TEs classified as cognates and non-cognates by two/three raters according to their phonological similarity
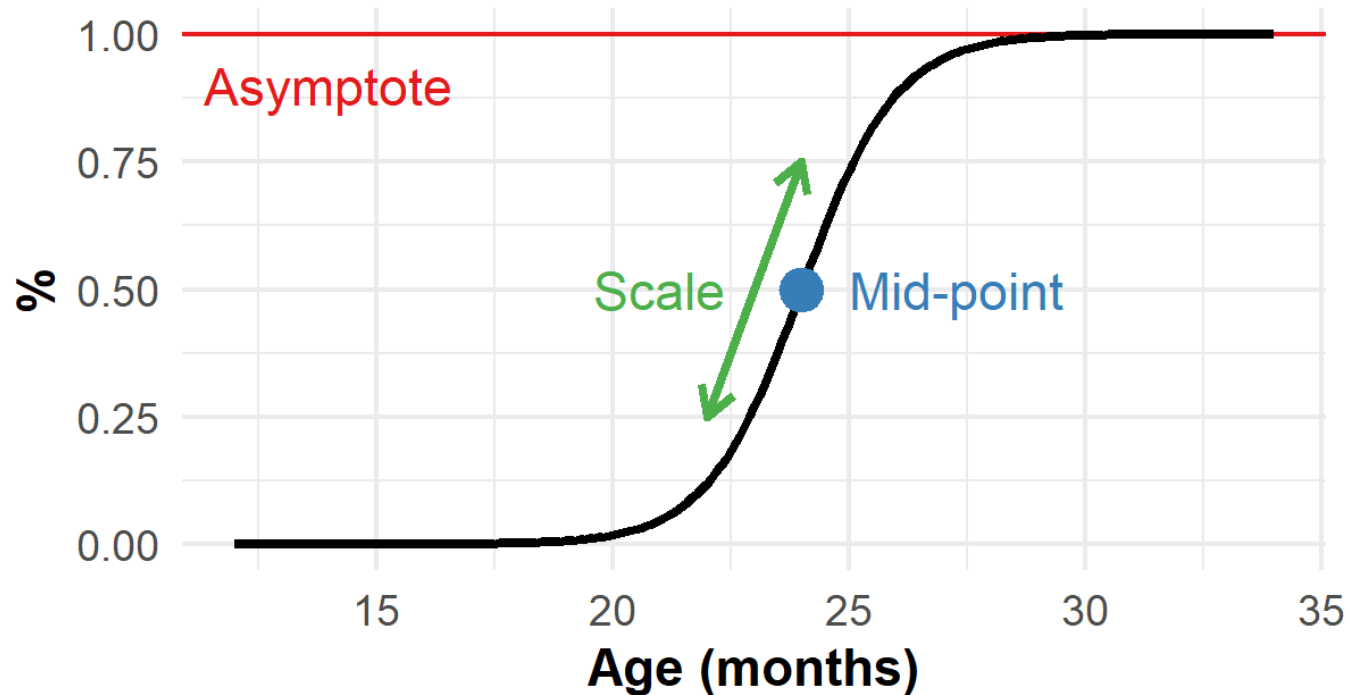
Dominance × Cognateness

Outcome
Proportion of infants that understan/produce each TE at different age bins.

Data analysis: Logistic curves

Parameters are **theoretically meaningful** (Mayor & Plunkett, 2011; Mahr et al., 2020)

Growth rate relatively stable until a "**sensitive age**" is reached until it *plateaus*

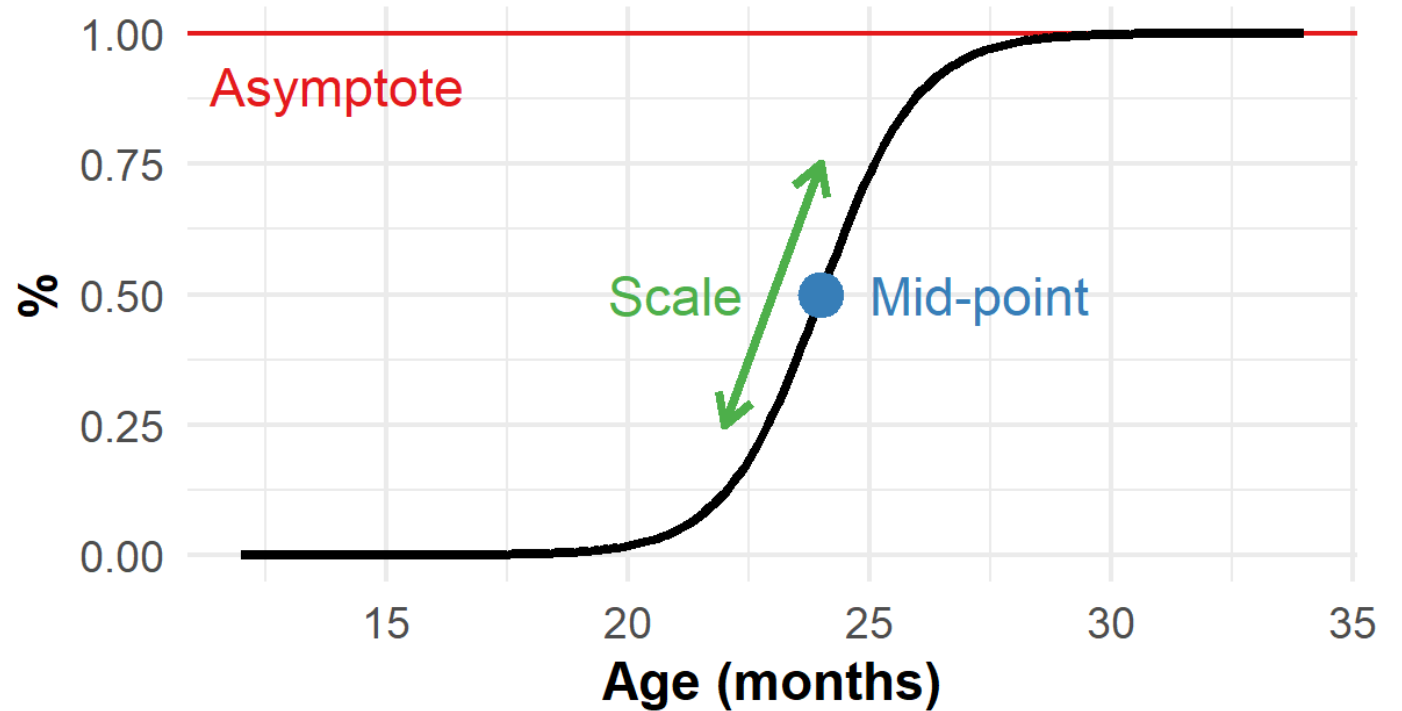Allows **explicit estimation** of its three parameters

Data analysis

$$y = \frac{Asymptote}{1 + e^{\frac{Mid-Age}{Scale}}}$$

$$Asymptote = \beta_0 + \varepsilon$$

$$Scale = \beta_0 + \varepsilon$$

$$Mid = \beta_0 + \beta_{0i} + (\beta_1 + \beta_{1i})Dominance + \cdots +$$
$$\beta_2 Cognateness + \beta_3 Dominance \times Cognateness + \varepsilon$$

Where i is a Translation Equivalent

Results: production

The model that includes **all main effects and interaction** is the one that fits the data the best (even correcting for the number of predictors included)

Model	df	AIC	BIC	Log. Likelihood	Likelihood ratio	p
Intercept	5	52,106.31	52,152.84	−26,048.15	-	
... + Dominance (L2/L1)	8	50,207.47	50,281.93	−25,095.74	1,904.84	< .001
... + Cognateness (NC/C)	9	50,205.36	50,289.12	−25,093.68	4.11	.043
... + Dominance × Cognateness	10	50,139.87	50,232.94	−25,059.93	67.49	< .001

Results: comprehension
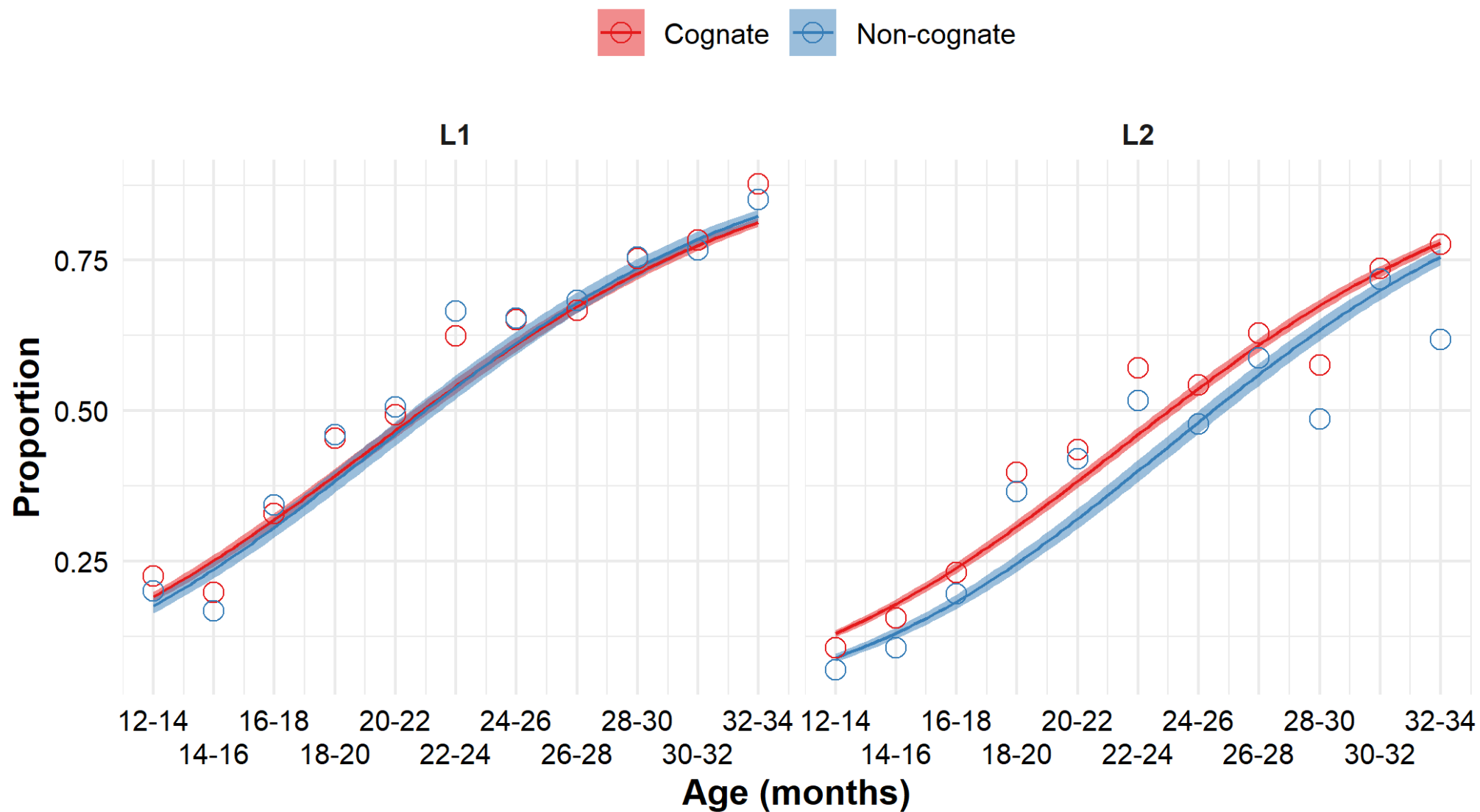
Results: comprehension

Term	Estimate	SE	Den. df	F	p	95% CI
Asymptote (max. %)						
Intercept	0.93	0.00	80,765	234.26	< .001	[0.92, 0.93]
Mid-point (age of acquisition)						
Intercept	4.75	0.14	80,765	33.11	< .001	[4.47, 5.04]
Dominance	−1.42	0.04	80,765	−33.75	< .001	[-1.5, -1.33]
Cognateness	−0.11	0.22	80,765	−0.49	.624	[-0.54, 0.33]
Dominance × Cognateness	0.70	0.08	80,765	8.51	< .001	[0.54, 0.86]
Scale (Steepness)						
Intercept	2.09	0.02	80,765	86.32	< .001	[2.05, 2.14]

Mid-point re-scaled parameters:
Intercept: 21.5 months
Dominance: -2.84 months
*Cognateness*: -0.22 months
Dominance × Cognateness: 1.40 months

Results: production

The model that includes **all main effects and interaction** is the one that fits the data the best (even correcting for the number of predictors included)

Model	df	AIC	BIC	Log. Likelihood	Likelihood ratio	p
Intercept	5	22,130.63	22,176.83	−11,060.31	-	
... + Dominance (L2/L1)	8	20,073.75	20,147.67	−10,028.88	2,062.87	< .001
... + Cognateness (NC/C)	9	19,884.72	19,967.88	−9,933.36	191.03	< .001
... + Dominance × Cognateness	10	19,853.16	19,945.56	−9,916.58	33.57	< .001

Results: production
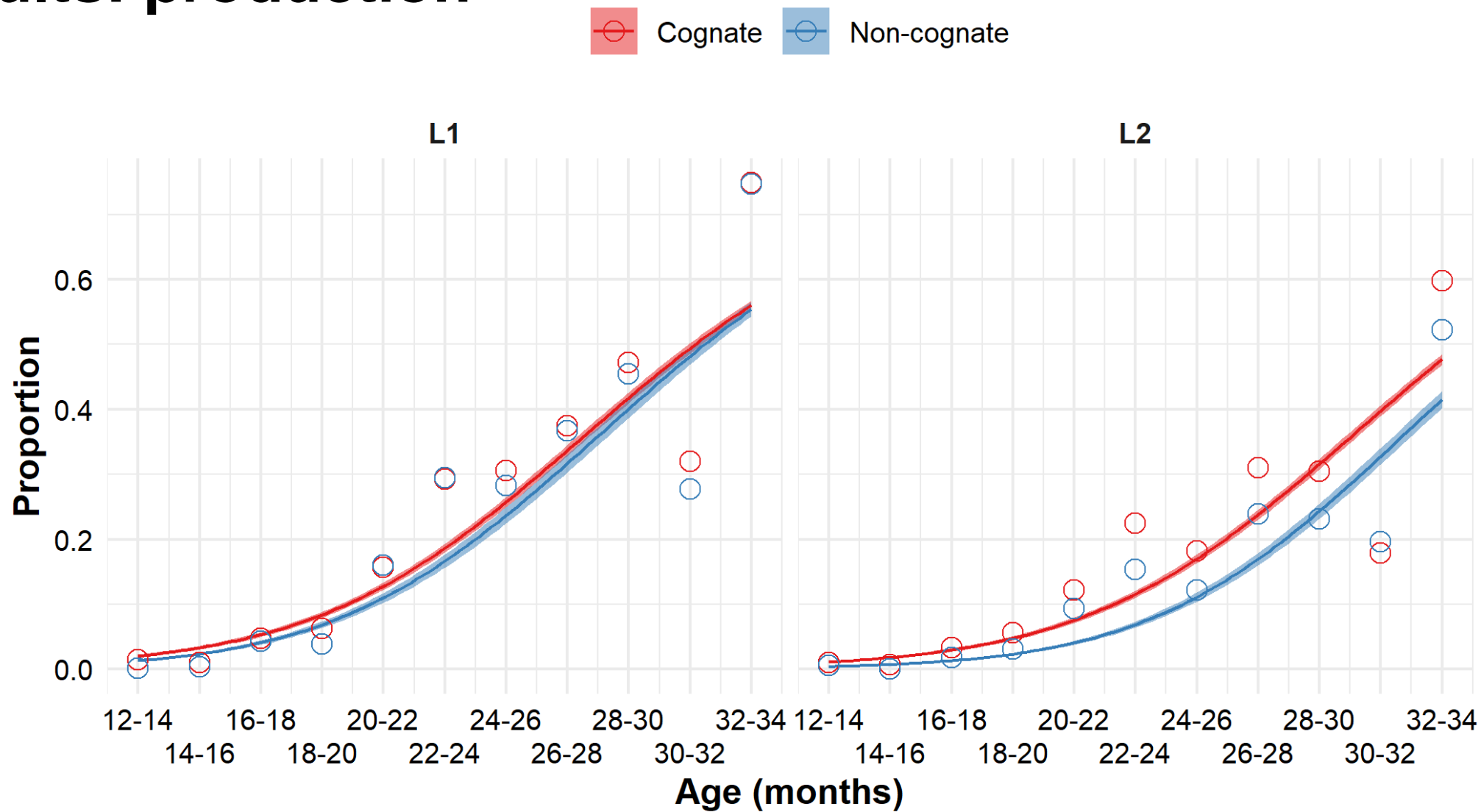
Results: production

Term	Estimate	SE	Den. *df*	F	p	95% *CI*
Asymptote (max. %)						
Intercept	0.73	0.01	75,487	107.81	< .001	[0.72, 0.74]
Mid-point (age of acquisition)						
Intercept	8.19	0.11	75,487	74.73	< .001	[7.97, 8.4]
Dominance	−1.54	0.05	75,487	−28.99	< .001	[-1.64, -1.43]
Cognateness	−0.34	0.17	75,487	−1.96	.051	[-0.68, 0]
Dominance × Cognateness	0.62	0.10	75,487	5.89	< .001	[0.41, 0.82]
Scale (Steepness)						
Intercept	1.66	0.02	75,487	69.12	< .001	[1.61, 1.71]

Mid-point re-scaled parameters:
Intercept: 28.38 months
Dominance: -3.08 months
Cognateness: -0.68 months
Dominance × Cognateness: 1.24 months

Results: predicted mid-points (AoA)
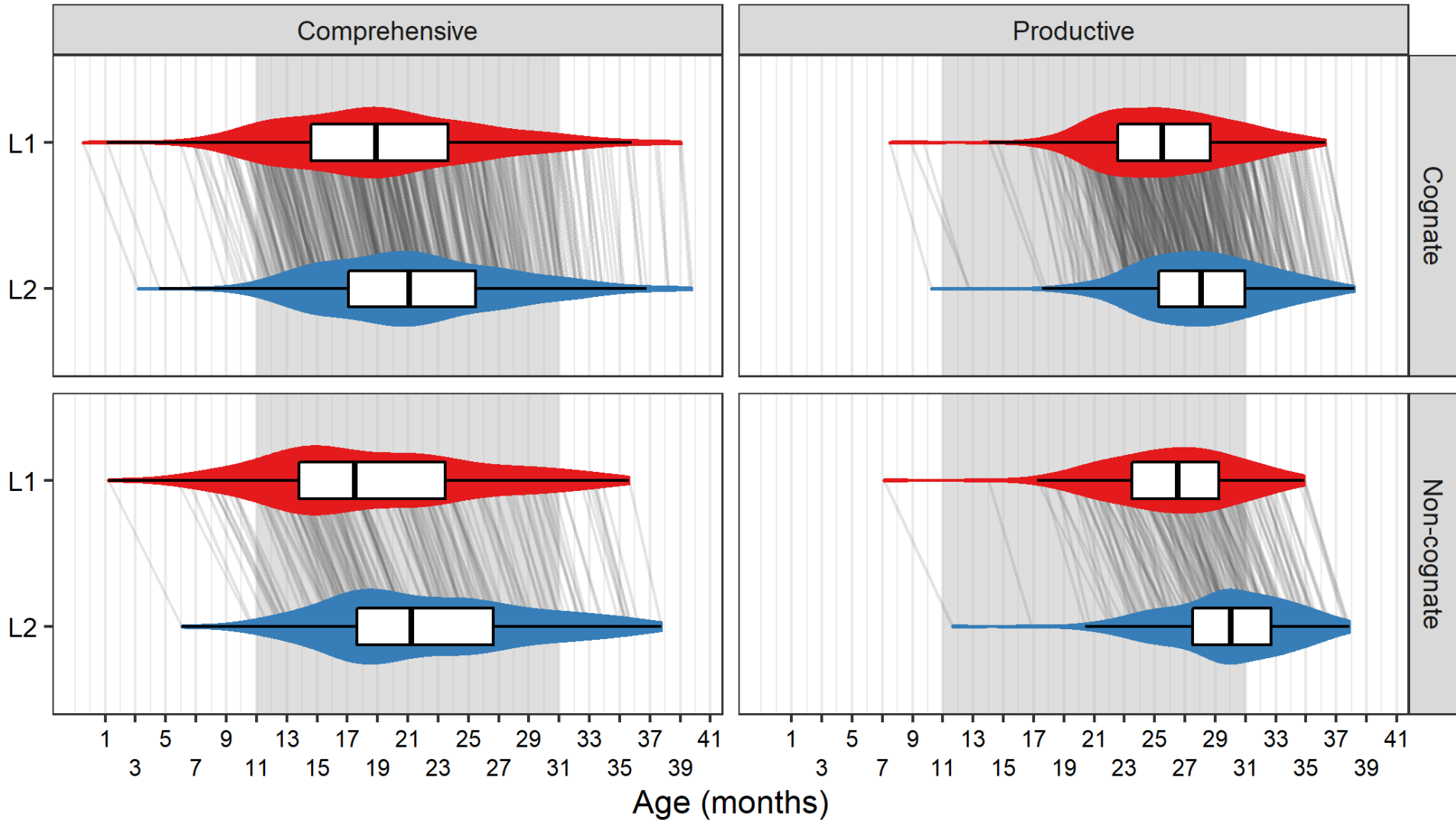
Discussion: Summary

- During early lexical acquisition, **cognates are acquired earlier** than non-cognate by Catalan-Spanish bilinguals

- This effect is (mostly) present in the **language of least exposure (L2)**, and in **production**

- **Shorter time difference** between the acquisition of members of **cognate TEs** than of non-cognate TEs

- Results consistent with the **parallel activation account** of bilingual lexical acquisition

- **Next steps**: more diversity of language pairs, wider age range, ideally longitudinal

- **Experimental** evidence needed

Discussion: Weak points

- **Cross-sectional data**: developmental conclusions sould be taken with care

- **Age range** of observed data falls a bit short

- *N* non-cognates very low relative to *N* cognates

- **Delay in production**? Psychometric artifact: parents of >16 mo children are not asked to choose between comprehension *or* comprehension and production in the classsic CDI. More likely to report production?

Lexical frequency

Questionnaire and item properties

Participants were randomly allocated on of the the four lists, and completed both the Catalan and the Spanish versions.

List	Frequency (Non-cognates)					Frequency (Cognates)				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Catalan										
List A	99	4.60	1.06	2.03	7.44	223	4.47	0.96	1.25	7.52
List B	93	4.39	0.98	1.94	6.94	225	4.49	0.90	0.86	7.52
List C	104	4.40	1.00	1.90	6.94	211	4.46	0.97	1.67	7.52
List D	102	4.45	0.94	2.10	6.94	210	4.44	0.85	1.51	7.52
Spanish										
List A	94	4.47	0.99	2.28	7.53	216	4.49	0.99	1.68	7.12
List B	92	4.44	0.93	2.16	7.53	236	4.35	0.96	1.68	7.22
List C	101	4.58	0.99	2.87	7.53	218	4.55	1.02	1.68	7.52
List D	104	4.54	1.05	2.70	7.53	220	4.41	1.00	1.68	7.32